ABSTRACT

SAS/GIS software displays, analyzes and queries spatial data. This paper discusses the GIS aspects of a larger application that incorporates several SAS products. It discusses which data was used for this application, importing the data, and setting up the SAS/GIS maps. Finally, this paper discusses how the maps were combined with reports, presentation graphs, images, and interactive visualization tools needed to analyze the attribute data in a complete application.

INTRODUCTION

To show one of the many possible uses of SAS/GIS software, an imaginary up-scale grocery store, Nature's Harvest, was postulated. Nature's Harvest specializes in selling:

- locally and organically grown fresh fruits and vegetables
- health food
- locally-made baked goods, pasta, jam
- low-fat, no-salt healthy snacks
- mineral water, fruit juices
- all-natural soaps and lotions
- healthy eating and fitness magazines

Nature's Harvest has stores in Charlotte (Mecklenburg County), North Carolina and would like to expand to another area in North Carolina. Raleigh is the next largest city and has enjoyed strong growth and good press recently, so the Wake County area is being considered. Before making this big decision, the factors that make the existing area need to be examined and compared to the proposed area. Other areas in North Carolina also need to be checked to make sure Wake County is the proper choice.

While the approach documented here may be somewhat simplified, it does represent a type of real world problem which can be tackled with SAS/GIS software.

DETERMINING THE DATA

There are two types of data used in SAS/GIS software, spatial data and attribute data. You need attribute data that matches your spatial data. While the most detailed data is usually best, you must balance cost, size and speed against more detail. For Nature's Harvest we started with a North Carolina county map that could be imported from the existing SAS/GMAPS.COUNTIES data set. To provide spatial data with further details for each county, we used the TIGER/Line Files for North Carolina. The TIGER files gave us street centerlines for a visual reference and census geography (like census tracts) to match with attribute data. We started with the census STF1A data for North Carolina which provided the 1990 demographic attribute data which included:

1. population
2. number of households
3. marital status
4. age groups
5. median home values

To get more up-to-date data as well as data more directly related to the up-scale grocery business, we obtained further attribute data from CLARITAS® at the county level for all of North Carolina. We also ordered census tract level data for Mecklenburg and Wake counties at the same time. This allowed us to examine the census tracts near our existing stores, build a model, and predict sales in various Wake County areas. We received five types of attribute data from CLARITAS:

1. Demographic Data
2. Lifestyle Data
3. Grocery Data
4. Magazine Data
5. PRIZM® Clusters

Each type of data is described in more detail below.

Demographic Data

The demographic data included the 1993 estimates for population and various household characteristics, like the number of households with:

- income > $50,000
- income > $50,000 and people age 25-34
- income > $50,000 and people age 35-44
- income > $50,000 and people age 45-54
- people who went to college
- children

Lifestyle Data

The lifestyle data contains an indexed value indicating the number of people engaged in various active hobbies. A value of 100 indicates typical activity for North Carolina. The following activities were included:

- bicycling
- jogging or running
- racketball
- golf
- tennis
- belongs to a health club
Grocery Data

The grocery data contains estimates on the total amount of money spent on groceries as well as the following categories:

- fresh fruits
- fresh vegetables
- non-frozen juices
- pasta and cereals

Magazine Data

The magazine data contains an indexed value indicating the number of people subscribing to various food related magazines. A value of 100 indicates typical number of subscribers for North Carolina. The following magazines were included:

- Bon Appetit
- Gourmet
- American Health
- Food and Wine

PRIZM Data

PRIZM is a market segmentation system that defines our areas according to distinct types or clusters. Each PRIZM cluster combines detailed demographics with product, media, and lifestyle preferences to create an accurate portrait of people in these areas.

Spatial Data

The spatial data for North Carolina counties was imported from the SAS/GRAPH maps. We choose census tracts for our further detail so that the spatial data could be imported from the Census Bureau TIGER/Line files.

Competitive grocery store locations in Mecklenburg and Wake counties were also obtained from CLARITAS. This allowed us to view our competitors along with the store size (in square feet), store name, and approximate sales volume. This data was imported as generic point data and then combined with the appropriate county data.

IMPORTING THE DATA

The North Carolina county map was imported from a subset of the SAS/GRAPH county map. The subset was created with the following data step and was then imported in the SAS/GIS import window as shown in Display 1. The imported map is shown in figure 1.

```sas
data NC;
  set MAPS.COUNTIES;
  where State = 37; /* FIPS code for NC */
run;
```

Display 1 Importing North Carolina Map

Figure 1 North Carolina County Map

SETTING UP THE GIS MAPS

We setup three GIS maps, one for North Carolina and one each for Mecklenburg and Wake County. Layers, actions, and labels were defined for each map.

Layers

SAS/GIS software displays three types of features: points, lines and areas. A layer is a logical group of features and a map has one or more layers. If a feature is not in a layer, you will not see it on your map. The North Carolina map has an area layer for the counties and a point layer for the major cities. The county layer was given a theme so that the counties were shaded according to the number of households with income greater than $50,000.

The Mecklenburg county map has the following layers:

- County: the county outline
- Tract: the census tract areas
- Street: the street lines are used for orientation. There were too many streets to effectively display when looking at the whole county, so this layer was defined to be visible after zooming in on the map.
- Harvest: points for the Nature's Harvest stores
- Groc: points for other grocery stores had associated attribute data estimating the yearly sales and the size of the store.
The Wake county map has the following layers:

**County**
the county outline was included to give a well-defined visual border to our study area.

**Tract**
the census tract areas are associated with the attribute data from CLARITAS and the census bureau.

**Street**
the street lines are used for orientation. There were too many streets to effectively display when looking at the whole county, so this layer was defined to be visible after zooming in on the map.

**Real**
points for the available real estate being considered for a new Nature's Harvest store has associated attribute data describing the property.

**Groc**
points for other grocery stores have associated attribute data estimating the yearly sales and the size of the store.

---

**Image**
Displays an image associated with a feature. The name of the image to display is stored in an attribute data set.

**Spatial**
Displays spatial data that can be used for further selections or edited.

The most basic type of action is a browse action. This allows you to view attribute data in a FSBROWSE window. FSBROWSE is a part of SAS/FSP® software that shows one observation from a data set in a user-customizable window. In a browse action, the observation shown corresponds to the feature selected on the map. In Display 2 the definition to browse the CLARITAS data for a selected county is shown.

---

**Actions**
Actions in SAS/GIS software allow you to interact with the attribute and spatial data. The types of actions are:

**Browse**
Uses FSBROWSE to display attribute data in a FSBROWSE window one observation at a time. The layout of the window can be customized using FSBROWSE.

**View**
Uses FSVIEW to display attribute data in a FSVIEW window. Many observations are visible at one time.

**Data**
Saves a subset of attribute data in a SAS data set.

**Program**
Saves a subset of attribute data in a SAS data set and runs a SAS program. This allows you to use other parts of the SAS System to analyze and generate reports from SAS/GIS software.

**Drill-down**
Loads another map into the SAS/GIS map window.

---

**Display 2** *Browse Action Definition*
A program action is the most powerful type of action. The program action defined in Display 3 subsets the CLARITAS and census data for the selected counties and then submits the program that generates the bar chart seen in figure 3.

---

**Display 3** *Program Action Definition*
Since city names are more familiar than county names, the city points and names were added to the North Carolina county map. The map was too crowded with all of the city names displayed, so most of the name labels were set to appear after the map display scale reaches 30 km/cm (see Display 4). The map display scale changes when you grow the window or zoom the map. So zooming in to an area will display all of the city names in that area.

The county name was added as a simple label to the Wake and Mecklenburg maps, so that you could more easily recognize which map you were working with.

The county name was added as a simple label to the Wake and Mecklenburg maps, so that you could more easily recognize which map you were working with.

Figure 3  County Lifestyle Comparison

Labels

Since city names are more familiar than county names, the city points and names were added to the North Carolina county map. The map was too crowded with all of the city names displayed, so most of the name labels were set to appear after the map display scale reaches 30 km/cm (see Display 4). The map display scale changes when you grow the window or zoom the map. So zooming in to an area will display all of the city names in that area.

The county name was added as a simple label to the Wake and Mecklenburg maps, so that you could more easily recognize which map you were working with.

The county name was added as a simple label to the Wake and Mecklenburg maps, so that you could more easily recognize which map you were working with.

EXPLORING THE DATA

Building a GIS application is an iterative process. Once some basic maps were created the data was explored to find other information that we wanted to display. A statistical model was built from the CLARITAS attribute data for Mecklenburg county census tracts using principal component analysis in SAS/INSIGHT® software. During this modeling, the most significant variable contributing to the Nature's Harvest sales turned out to be the number of households with an income greater than $50,000.

While exploring the North Carolina county map, we saw that Wake County had the next highest number of households with an income greater than $50,000 (see Figure 2). In fact Wake had almost twice as many of these households as the next ranking county, Guilford. This helped reinforce the original idea of locating in Wake County. We developed some graphical comparisons of the counties to show other factors favoring Wake County as well. Figure 3 shows a comparison of the lifestyle indices for the three counties. Wake showed a stronger participation than Mecklenburg in every category while Guilford showed less participation. Figure 4 showed the similarity in demographics between Mecklenburg and Wake while Figure 5 showed that there were more members of the "yuppie" PRIZM cluster in Wake County. All of this reinforced the decision to locate in Wake County.
BUILDING AN APPLICATION

To combine the maps used above into an integrated application to be used in reviewing the selection process, several actions were added. A drill-down action was added to the county map (see Display 5) to allow the Mecklenburg and Wake County maps to be accessed directly. The following data step creates the data needed to drill-down to the appropriate map for the selected county.

```sql
data HARVEST.NCMAPS;
    length Map $ 26;
    input County Map;
    cards;
135 HARVEST.MAPS.MECK
189 HARVEST.MAPS.WAKE
run;
```

Display 5  Drill-Down Action Definition

With this action you can select either Wake or Mecklenburg county and perform the DRILL action to bring up the respective map with further details.

On a tour of the competing stores, photographs were taken and scanned in to be used in an image action. This image action was added to the Wake County map so that an image of the front of each grocery store would be displayed when it was selected. When you are exploring the data, these images remind you of the stores seen on your tour.

A program action with a simple report showing the total population, households, households with income greater than $50,000 and predicted sales was also added. This allows you to select a group of tracts around a potential site and see a summary of the tracts in that area.
CONCLUSION

SAS/GIS software can be a valuable part of an application. It provides a place to integrate the display, selection, and analysis of various data. The process of matching attribute and spatial data also helps define what data is needed. From this application we were able to determine that Wake County was a good choice for locating our new store. Now that the data is organized we can easily investigate within Wake County for the best site.

When an application uses large amounts of data, a picture is worth a thousand words. And when that data is spatially related, a map is worth a thousand pictures.

ACKNOWLEDGEMENTS

We would like to thank CLARITAS and Geographic Data Technologies for providing data for this application and the other people at the Institute who spent time on this project. For putting together this application: Lina Land, Zhongqiang Liu, Jade Walker, Bron Bailey and Mark Brown. For working on SAS/GIS software: Dave Jeffreys and Jeff Phillips.

SAS, SAS/FSF, SAS/GRAPH, SAS/GIS and SAS/INSIGHT are registered trademarks or trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.